Author Names & Affiliations

- Ralph Roskies - University of Pittsburgh
- Nancy Wilkins-Diehr - San Diego Supercomputer Center
- Lonnie Crosby - U. of Tennessee Knoxville
- Sergiu Sanielevici - Pittsburgh Supercomputing Center
- John Cazes - Texas Advanced Computing Center
- Marlon Pierce - Indiana University
- Jay Alameda - National Center for Supercomputing Applications

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Computational science

Title of Submission

People: A fundamental but often overlooked component of an impactful cyberinfrastructure

Abstract (maximum ~200 words).

There has been a tremendous growth in the science and engineering fields using advanced cyberinfrastructure (CI) and in the technologies that underlie it. As a result, the skills and knowledge needed to apply this advanced CI to solve research problems becomes even more demanding. From complex multi-core processors with special features to parallel programming, data analytics, cloud computing, workflows, visualization and data mining, the expertise needed to tackle even a single research challenge continues to grow. Critical to these scientific advances are the "bridge people", talented staff with enough domain expertise to understand the research goals and enough technical expertise to be able to apply the right solutions. Programs that provide career opportunities and training of such experts (both at leading centers and on campuses throughout the country) have significantly amplified investments in CI in the past and must continue to do so going forward.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

It is now a commonplace observation that computing is ubiquitous, not only in traditional science and engineering research but in many other disciplines as well - humanities, economics, sociology, geographic information systems, genomics and more. These fields expand

and change rapidly as does the computing technology they utilize. No longer is an understanding of parallelization and visualization sufficient to advance computational science. Today there are architecturally distinct technologies (multi-core and manycore processors, GPUs, and processors with high bandwidth memory such as KNL), and distinct programming styles (Hadoop, Spark, virtual machines, containers, and workflows). There are also software technologies and approaches that are favored among particular research fields, such as large scale data analytics and machine learning. It is likely that the future will require more and more specialized skills.

The challenge is that there is a mismatch between the expertise needed to be at the forefront of science and engineering research and the expertise needed to make best use of the technical advanced cyberinfrastructure (CI). We have learned in our decades of providing leading-edge advanced resources to the national research community that effectively using such technologies to facilitate scientific progress requires a cadre of talented staff, professionals whom we refer to as CI experts. These professionals combine familiarity with several fields of computational science and engineering with a deep knowledge of underlying computer systems and of the design and implementation principles, tips, and tricks for optimally mapping scientific problems, codes, and middleware to these resources. This group must include experts not only in the traditional numerical applications of advanced computing systems but also in extreme digital data management, workflow engineering, and the creation and maintenance of scientific gateways.

The payoff of connecting disciplinary leaders with this skilled set of CI experts results in increased code performance (sometimes by factors of 100 or more) or enables researchers to architect solutions using multiple technologies that they would not have otherwise undertaken. These types of gains have been achieved in XSEDE's Extended Collaborative Support Service (ECSS) program. Speedups of this magnitude mean that a researcher has results in hours rather than weeks. Implementation of these solutions can have transformative effects on the research these disciplinary leaders pursue. Both significant speedups and the incorporation of utterly new technological approaches can completely change the nature of the scientific inquiry, open the eyes of researchers to possibilities they had not considered and transform the types of problems they can tackle. The result is a major impact on scientists' productivity.

CI experts can also significantly reduce the barrier of entry to the leading-edge computing systems to communities previously unaware of the power of advanced CI. Improvements they make to widely used community codes frees up available computing time and reduces execution time for all researchers running that code. There are many examples of these successes in XSEDE's Extended Collaborative Support Service (ECSS) program and also on campuses via the Campus Champion and ACI-REF (Advanced CI Research and Education) programs.

Although many CI experts have had traditional training, education, and research experience, their CI expertise over time has enabled them to contribute to research projects and support activities outside the realm of their formal education and training. CI experts are well suited to not only to apply CI technologies to the research fields in which they were trained, but to contribute new, multidisciplinary ideas and approaches to disparate research areas.

The challenge is that we have not yet found a sustainable way to produce and retain these CI experts.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

A critical component in making effective use of cyberinfrastructure is well-trained CI experts. The CI expert is often the avenue by which science and engineering research derives its greatest impact from CI. In this "bridge" role, the CI expert not only enables the utilization of CI resources but also disseminates its impact within the research community. In order to perform this role effectively CI experts must have different skills, knowledge, and experience. They need to be able to speak the language of the researchers, oftentimes across multiple domains. They must stay current with developments in CI technologies, including software and hardware. While CI experts usually have training and education in a science and engineering discipline, they often have gained experience in various disciplines and CI technologies through extracurricular means. They are thus unique and valuable contributors to research, serving as an important bridge across disciplines.

It is imperative that these CI experts be appropriately involved in current science and engineering research. However, in many programs sponsored by the National Science Foundation (NSF) support for research activities tend to fall to faculty, students, or postdoctoral fellows. Future investments in CI should address the career, skills development, and research involvement of CI experts as fundamental to the progress of science and necessary for advanced CI to have a transformative impact on future science and engineering research.

Because of the diversity of expertise needed - both across disciplines and across technologies, programs which fund a critical mass of CI experts are a necessary component of the national cyberinfrastructure going forward.

Although many CI experts have an advanced degree in a scientific discipline, most are not pursuing active research in that discipline, and are therefore not viewed as professionals in that discipline. Neither are they viewed as professionals in computer science, since most computer science degree programs do not include an advanced degree in computational science, which is much closer to what these people do. They are thus not viewed as peers either by the research scientists or the traditional computer scientists. Thus there is a severe issue with their career path. Recognition that CI experts are important to the scientific landscape and instantiating that in proposal calls does help build career paths for these important people.

Given today's university structures, no single university is likely to produce CI experts who have the technical mastery for multi-core optimization, visualization, workflow development and an understanding of digital humanities, machine learning, and computational fluid dynamics and computational chemistry. Rather, each institution is likely to develop some CI experts in particular disciplines and specific technologies. Disparate groups must be brought together address such complexity.

Institutionally, Teragrid and XSEDE have shown that with the right ecosystem, these experts can function very effectively even though they are geographically distributed, and can respond to needs of a very wide set of disciplines utilizing very differing technologies. For example, the pooling of expertise has allowed ECSS to support experts in digital humanities and workflows, which it is unlikely that an individual center could have provided due to insufficient demand at any one site. That centralization also enables cross-pollination of knowledge between disciplines and resources. We have often seen advanced user support professionals transmitting advances and insights at the algorithmic, numerical, coding, and optimization levels between fields of application and between computing systems.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Major considerations that need to be addressed include workforce development and career stability. As mentioned, most computer science programs do not have a degree program for these CI experts. Nor do the disciplinary degree programs. Curricular reform is part of the solution, but is often slow to evolve. Proactive apprentice or internship programs for students, making them aware of careers they probably don't know exist, is critically important.

An additional challenge in the sciences is that these CI experts are much sought after by industry, such as Microsoft, Google, Facebook, or Wall Street. One might think that market mechanisms would lead universities to continue to produce such CI experts. But university curricula adapt very slowly to market mechanisms. NSF could materially affect the rate of production of CI experts by making clear that the continued production of CI experts is a national imperative.

It is essential that we solve the CI expert pipeline problem at all levels and from all demographic and socio-economic sources. We must provide the multidisciplinary education, training, and mentoring for students, culminating in capstone projects, theses, internships, apprenticeships etc., that make significant contributions to specific CI research efforts. Some of these students will be the next generation of CI experts; a focus on underserved and underrepresented students will tap a large new reservoir of talent. We need to have a fundamentally better understanding of the incentives that keep the current cohort of CI experts in place rather than in higher paying positions in industry. Finally, we need to have continuing education and training programs to attract "encore career" CI experts.

Consent Statement